Integrating Multimodal MRIs for Adult ADHD Identification with Heterogeneous Graph Attention Convolutional Network

Dongren Yao[1,2], Erkun Yang[1], Li Sun[3], Jing Sui[4(✉)], and Mingxia Liu[1(✉)]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
mxliu@med.unc.edu

[2] Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[3] National Clinical Research Center for Mental Disorders and Key Laboratory of Mental Health, Ministry of Health, Peking University, Beijing 100191, China

[4] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100678, China
jsui@bnu.edu.cn

Abstract. Adult attention-deficit/hyperactivity disorder (ADHD) is a mental health disorder whose symptoms would change over time. Compared with subjective clinical diagnosis, objective neuroimaging biomarkers help us better understand the mechanism of brain between patients with brain disorders and age-matched healthy controls. In particular, different magnetic resonance imaging (MRI) techniques can depict the brain with complementary structural or functional information. Thus, effectively integrating multi-modal MRIs for ADHD identification has attracted increasing interest. Graph convolutional networks (GCNs) have been applied to model brain structural/functional connectivity patterns to discriminate mental disorder from healthy controls. However, existing studies usually focus on a specific type of MRI, and therefore cannot well handle heterogeneous multimodal MRIs. In this paper, we propose a heterogeneous graph attention convolutional network (HGACN) for ADHD identification, by integrating resting-state functional MRI (fMRI) and diffusion MRI (dMRI) for a comprehensive description of the brain. In the proposed HGACN, we first extract features from multimodal MRI, including functional connectivity for fMRI and fractional anisotropy for dMRI. We then integrate these features into a heterogeneous brain network via different types of metapaths, with each type of metapatch corresponding to a specific modality or functional/structural relationship between regions of interest. We leverage both intra-metapth and inter-metapath attention to learn useful graph embeddings/representations of multimodal MRIs and finally predict subject's category label using these embeddings. Experimental results on 110 adult ADHD patients and 77 age-matched HCs suggest that our HGACN outperforms several state-of-the-art methods in ADHD identification based on resting-state functional MRI and diffusion MRI.

Keywords: Attention-deficit/hyperactivity disorder · Graph convolution network · Resting-state functional MRI · Diffusion MRI

I. Rekik et al. (Eds.): PRIME 2021, LNCS 12928, pp. 157–167, 2021.
https://doi.org/10.1007/978-3-030-87602-9_15

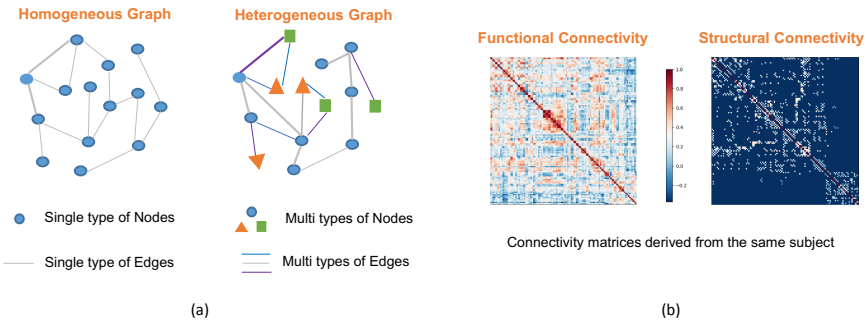

Fig. 1. Illustration of (a) homogeneous and heterogeneous graphs, and (b) functional and structural connectivity networks derived from resting-state fMRI and diffusion MRI of the same subject.

1 Introduction

Attention-Deficit/Hyperactivity Disorder (ADHD) is one of the most common neurodevelopmental disorders that occur in children, adolescents, and adults. Patients with ADHD often exhibit an ongoing pattern of inattention, hyperactivity and impulsive behavior [1].

Previous studies usually pay more attention to children with ADHD and have made substantial achievements in computer-aided diagnosis [2]. However, ADHD can last into adulthood, and symptoms along with adult patients can lead to difficulty at home or work and in social settings. The current diagnosis of ADHD in adults gathering multiple resources is mainly based on subjective observation or rating scale, such as diagnostic manual Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). Considering the fact that ADHD-associated symptoms would change over time, it is essential to employ objective imaging biomarkers to help us better understand the brain mechanism of adult ADHD for timely intervention and effective treatment [3–6].

In the past two decades, the use of non-invasive neuroimaging techniques such as electroencephalography (EEG), magnetic resonance imaging (MRI) and functional MRI to describe and understand the brain has progressed rapidly [7–9].

With the development of these brain imaging technologies, we have more direct ways to observe and measure those brain structural and functional alterations between healthy control and pathological conditions [10–13]. Currently, the etiological bases and neural substrates of adult ADHD are far from being fully understood. For instance, patients with ADHD suffer from inattention, hyperactivity, and impulsive behaviors that are not matched with their own age.

To better understand the neurobiological mechanisms underpinning this mental disorder, many fMRI-based studies have reported that large-scale brain dysfunctions may lie in ADHD patients [14,15]. Also, diffusion tensor imaging (DTI) based studies [16–18] have suggested that the underdevelopment of gray or white

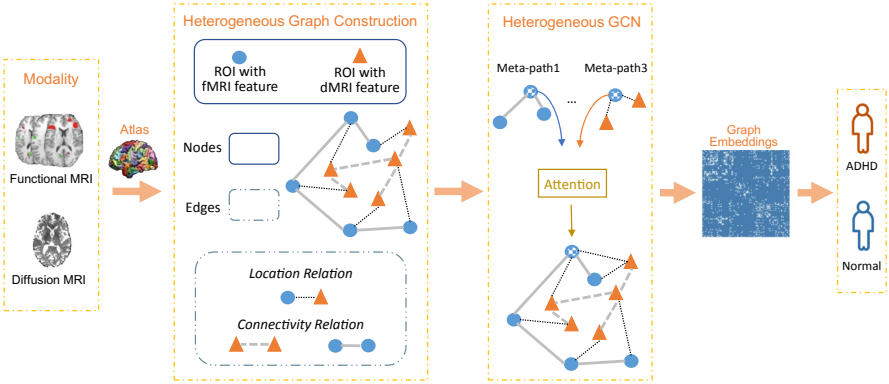

Fig. 2. Illustration of our Heterogeneous Graph Attention Convolutional Network (HGACN) for ADHD identification with multimodal MRI. The whole framework contains 3 parts: 1) Based on preprocessed functional and diffusion MRIs and specific atlas, we calculate Functional Connectivity (FC) from fMRI and Fractional Anisotropy (FA) from dMRI. 2) Based on the calculated functional and structural connectivity matrix, we construct a heterogeneous graph with two different nodes and three kinds of edges for each subject. 3) Based on the constructed graph, we employ a heterogeneous graph convolution network to learn a new graph embedding for each subject. With learned graph embedding, a fully-connected layer is leveraged for predicting the category label of a subject.

matter structure in the frontal lobe, striatum, and thalamus area may lead to the emergence of ADHD when they were in childhood. However, most of the existing studies focus on using a single MRI modality (*e.g.*, fMRI or dMRI) for investigating ADHD. Considering that these modalities can provide multi-view diverse and complementary information, it is highly desired to design a unified framework to effectively leverage multimodal MRIs for ADHD diagnosis.

Apart from advances in multimodal neuroimaging techniques, the development of graph convolution networks (GCNs) has allowed researchers to treat the human brain's structural and functional connectivity patterns as the "connectome" to learn rich graph embeddings/representations. Many previous works use graph theory information such as modularity and hierarchy. Based on a single modality, several previous studies [9,19–21] employed Kipf or Chebyshev GCN and metric learning framework to measure the similarity between patients and healthy controls. With multiple modalities, Xing *et al.* proposed a DS-GCN model, consisting of GCN and Long term short Memory (LSTM) based on structural MRI (sMRI) and fMRI data, for automated diagnosis of mild cognitive decline [22]. Several studies employed GCN models to perform brain disorder analysis using DTI and fMRI data [23,24]. These studies have shown that using GCN can achieve superior performance than traditional learning-based methods on brain image analysis. However, they generally treat graphs corresponding

to different modalities as homogeneous graphs, ignoring inherent inter-modality heterogeneity of multimodal MRIs. As shown in Fig. 1 (a), compared with homogeneous graphs, multi-type nodes and edges comprise heterogeneous graphs, containing richer information. Current GCN models dealing with multimodal data (no matter with early or late fusion strategy) mine each modality separately, thus cannot effectively integrate multimodal data for subsequent analyses. As shown in Fig. 1 (b), after constructing heterogeneous fMRI-based and dMRI-based graphs (*i.e.*, connectivity matrices) for each subject, it is expected to design a unify framework to learn useful graph embeddings/representations for assisting in automated ADHD diagnosis.

In this paper, we propose a Heterogeneous Graph Attention Convolution Network (HGACN) to fuse structural and functional MRIs for ADHD identification. As shown in Fig. 2, we first extract diverse features based on functional MRI (fMRI) and diffusion MRI (dMRI) to depict the brain topology. For each subject, we employ the Automated Anatomical Labelling (AAL) template to generate a functional connectivity matrix/graph from resting-state fMRI and a fractional anisotropy (FA) matrix from dMRI. After pre-processing, we then model these features into one heterogeneous brain graph through different types of metapaths, with each type of meta-pathes corresponding to a specific modality or functional/structural relationship between paired regions of interest. We leverage both intra-metapth and inter-metapath attention to aggregate and learn useful graph embeddings for representing multimodal MRIs, and finally use them to identify adult ADHD from age-matched healthy controls (HCs). Experimental results on 110 adult ADHD and 77 age-matched HC subjects demonstrate the effectiveness of HGACN in integrating fMRI and dMRIs for ADHD identification. To the best of our knowledge, this is among the first attempts to fuse resting-state functional MRI and diffusion MRI data via an end-to-end heterogeneous GCN model for automated brain disorder analysis.

2 Materials and Method

2.1 Data and Image Pre-processing

The studied ADHD dataset contains 110 adult ADHD patients and 77 age-matched HCs recruited from a local hospital. Each subject is represented by three imaging modalities, including structural MRI (sMRI), rs-fMRI and dMRI. The demographic information about the studied subjects can be found in Table 1.

For rs-fMRI data, we leverage the Data Processing Assistant for Resting-State fMRI (DPARSF) to calculate functional connectivity matrix based on the AAL atlas. Specifically, we first discard the first ten time points, followed by slice timing correction, head motion correction, regression of nuisance co-variants of head motion parameters, white matter, and cerebrospinal fluid (CSF). Then, fMRI data are normalized with an EPI template in the MNI space, and resampled to the resolution of $3 \times 3 \times 3\,\mathrm{mm}^3$, followed by spatial smoothing using a 6 mm

Table 1. Demographic information of studied subjects in the Adult ADHD dataset. M: Male; F: Female; Std: Standard Deviation.

Category	#Subject	Site 1	Site 2	Gender (F/M)	Age (Mean±Std)
Adult ADHD	110	72	38	37/73	25.9 ± 4.9
Healthy Control (HC)	77	43	34	34/43	26.0 ± 3.9

full width half maximum Gaussian kernel. Finally, the AAL atlas, with 116 predefined regions-of-interest (ROIs) including cortical and subcortical areas, are nonlinearly aligned onto each scan to extract the mean time series for each ROI.

The dMRI data are preprocessed through FMRIB Software Library (FSL) with the following operations. 1) We first use the brain extraction tool (BET) to remove non-brain tissue; 2) then leverage nonuniform intensity normalization (N3) algorithm for filed bias correction on T1-weighted sMRIs; 3) further employ eddy current induced distortion to correct DWI scans; 4) also adjust the diffusion gradient strength accordingly, and 5) finally calculated the fractional anisotropy (FA) from DWI scans based on diffusion tensor fitting.

2.2 Method

As shown in Fig. 2, our HGACN model aims to combine multimodal MRIs into a unified heterogeneous graph learning framework for identifying ADHD patients from HCs. For each subject, we construct two heterogeneous graphs based on its functional and structural connectivity features derived from fMRI and dMRI, respectively. More details can be found in the following.

Construction of Heterogeneous Graph. The structure of a heterogeneous graph can be defined as $\mathbf{G} = (\boldsymbol{V}, \boldsymbol{E})$, where $\boldsymbol{V}$ and $\boldsymbol{E}$ denote the nodes and edges in the heterogeneous graph, respectively. The node set $\boldsymbol{V}$ has two different types of nodes (with each node corresponding to a specific ROI): 1) *functional nodes* $\boldsymbol{V}_f$, and 2) *structural nodes* $\boldsymbol{V}_d$. The edge set $\boldsymbol{E}$ consists of three types of pairwise ROI relationships: 1) *functionally connected edges* $\boldsymbol{E}_{ff}$ that represent functional uniformity between $\boldsymbol{V}_f$ nodes based on their Pearson correlation, 2) *structurally connected edges* $\boldsymbol{E}_{dd}$ that express structural connectivity coherence between $\boldsymbol{V}_d$ nodes based on their fractional anisotropy, and 3) *location-related edges* $\boldsymbol{E}_{df}$ that connect two different type of ROIs ($\boldsymbol{V}_d$ and $\boldsymbol{V}_f$) which are physically close in the predefined atlas. In this work, we employ AAL atlas containing 116 ROIs for brain ROI parcellation. Besides, for a heterogeneous graph, we define a metapath to describe a composite relationship ($\boldsymbol{V}_1 \xrightarrow{E_1} \boldsymbol{V}_2 \xrightarrow{E_2} \cdots \xrightarrow{E_l} \boldsymbol{V}_{l+1}$) between a set of nodes (*e.g.*, $\boldsymbol{V}_1$ and $\boldsymbol{V}_{l+1}$).

After constructing heterogeneous graphs, we propose a novel Heterogeneous Graph Attention Convolutional Network (HGACN) to learn new graph embeddings/representations for input subjects. Denote I as an identity matrix, A as an adjacency matrix, and M as a degree matrix. Similar to standard GCNs [9,25]

that are performed on homogeneous graphs $\mathcal{G}$, the aggregation operation with an adjacency matrix $A^{'} = A + I$ is defined as follows:

$$H^{(l+1)} = \sigma(\widetilde{A} \cdot H^{(l)} \cdot W^{(l)}) \tag{1}$$

where $\sigma(\cdot)$ is an activation function such as ReLU, $\widetilde{A} = M^{-\frac{1}{2}} A^{'} M^{-\frac{1}{2}}$ denotes the symmetric normalized adjacent matrix, and $W^{(l)}$ is trainable parameters for the l-th layer in GCN.

Through Eq. (1), we can only deal with nodes containing the same shared embedding space, therefore failing to handle heterogeneous graphs that contain different types of nodes and edges. To address this issue, we define a new graph convolution to explicitly take advantage of different types of node and edge information in a heterogeneous graph, as follows:

$$H^{(l+1)} = \sigma(\sum_{\tau \in \mathcal{T}} \widetilde{A}_{\tau} \cdot H_{\tau}^{(l)} \cdot W_{\tau}^{(l)}) \tag{2}$$

where $\mathcal{T}$ denotes all types of nodes ($\mathcal{T} = 2$ in this work) in heterogeneous graph, $\widetilde{A}_{\tau}$ denote the adjacency matrix corresponding to the τ-th type ($\tau = 1, \cdots, \mathcal{T}$). Based on Eq. (2), the new representation $H^{(l+1)}$ is generated by aggregating different types of edges with different transformation matrices $W_{\tau}^{(l)}$ from their neighboring nodes $H_{\tau}^{(l)}$.

Note that different types of neighboring nodes would have different impacts on a specific node, and the same type of nodes may be similarly affected by their neighbors. To model these characteristics, we propose to capture the intra- and inter-metapath attention information to learn more effective graph embeddings.

Intra-metapath Aggregation. For a specific node v, attention for intra-metapath aggregation will learn the weighs of the same types of neighboring nodes. Specifically, we denote $h_{\tau} = \sum_{v'} \widetilde{A}_{vv'} h_{v'}$ as the embedding of the τ-th type of nodes based on their neighbor node features $h_{v'}$. Then, we calculate the intra-metapath attention score as shown below:

$$\alpha_{\tau} = \sigma(\mu_{\tau}^{T} \cdot [h_v \oplus h_{\tau}]) \tag{3}$$

where μ_{τ} is the attention vector for the τ-th type τ and $\oplus$ denote the concatenation operation. After that, we compute the final node-level attention weight of the τ-th node type based on a softmax function as follows:

$$\alpha_{\tau} = \frac{\exp(\alpha_{\tau})}{\sum_{\tau' \in \mathcal{T}} \exp(\alpha_{\tau'})} \tag{4}$$

Inter-metapath Aggregation. Besides the intra-metapath aggregation on a specific node type, we also propose to model the importance of different neighbor nodes between different types. Denote N_v as a node set, with each node denoting

a neighbor of a center node v. Note that these neighbor nodes and the center node v may be defined in a different manner (*i.e.*, belonging to different types of node sets). Given a specific node v defined in the τ-th node type, we calculate the inter-metapath attention based on its neighbour nodes $v' \in N_v$ as follows:

$$\beta_{vv'} = \frac{\exp(\sigma(n^T \cdot \alpha_{\tau'}[h_v \oplus h_{v'}]))}{\sum_{i \in N_v} \exp(\sigma(n^T \cdot \alpha_i[h_v \oplus h_i]))} \tag{5}$$

where h_v and $h_{v'}$ denote the embeddings of the node v and the node v', n is a to-be-learned attention vector, and $\alpha_{\tau'}$ and α_i are calculated from Eq. (3).

Based on intra- and inter-metapath aggregation with two attention operations, the final heterogeneous graph convolution can be written as follows:

$$H^{(l+1)} = \sigma(\sum_{\tau \in \mathcal{T}} \mathcal{B}_\tau \cdot H_\tau^{(l)} \cdot W_\tau^{(l)}) \tag{6}$$

where the attention matrix $\mathcal{B}$ is a diagonal matrix whose v-th diagonal element is defined in Eq. (5).

Implementation. The whole HGACN framework stacks 3 heterogeneous graph convolution layers and one softmax layer for the final classification based on our learned new graph embedding for each subjects. We optimize the HGACN via the Adam algorithm, with the learning rate of 0.001, the number of epochs of 150, and the mini-batch size of 15. We implemented the HGACN based on Pytorch, and the model was trained with a single GPU (NVIDIA GeForce GTX TITAN with 12 GB memory).

3 Experiment

Experimental Setup. We evaluate the proposed HGACN and competing methods on the ADHD dataset through a 5-fold cross-validation strategy. The performance of MDD identification from age-matched HCs is measured by three metrics, including accuracy (ACC), sensitivity (SEN), and specificity (SPE).

Competing Method. We first compare the HGACN method with two conventional learning methods based on two connectivity matrices, including 1) Principal Component Analysis (PCA) + Support Vector Machine with Radial Basis Function kernel (**PS**), and 2) Clustering Coefficients (CC) with SVM (**CS**). The CC can measure the clustering degree of each node in a graph and can be treated as a feature selection algorithm. Hence, two different feature selection or extraction methods are used to reduce the dimensionality of original features. The number of components in PS is chosen from 3 to 50, with the step size of 1 via inner cross-validation. The CS method is associated with the degree of network sparsity, where the sparsity parameter is chosen from $\{0.10, 0.15, \cdots, 0.40\}$ according to cross-validation performance. The parameter C of SVM (with RBF kernel)

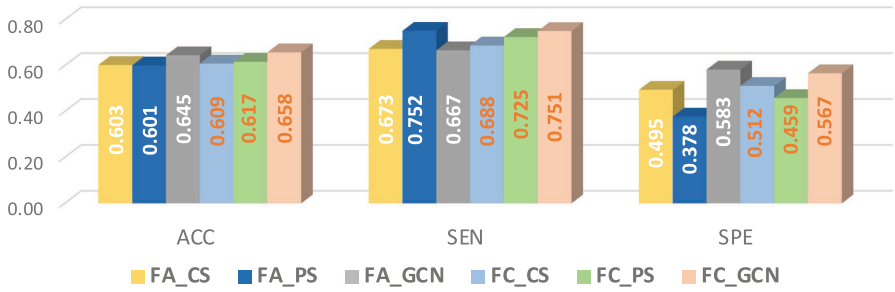

Fig. 3. Results of three methods (*i.e.*, CS, PS and GCN) with single modality data for ADHD vs. HC classification, *i.e.*, fractional anisotropy (FA) derived from dMRI and functional connectivity (FC) from resting-state fMRI.

used in CS and PS is chosen from $\{0.80, 0.85, \cdots, 3.00\}$ via cross-validation, and we use default values for the other parameters.

We also compare the proposed HGACN with the state-of-the-art deep learning method, *i.e.*, graph convolutional network (**GCN**) [25] with K-Nearest Neighbors graph generated from structural or functional connectivity matrix. The parameter k for constructing KNN graphs is chosen from $\{1, 2, \cdots, 30\}$. After modeling the node-centralized local topology (reflected by vertices and their connectivity) as the group-level shared by all subjects, the GCN method also contains 3 graph convolutional layers and one fully-connected layer.

Since two different data modalities are used in our study, we employ both early fusion (**EF**) and late fusion (**LF**) strategies for different methods to take advantage of these modalities. Specifically, for early fusion, we simply concatenate different types of features for the same subject as input features, followed by a specific machine learning algorithm (*e.g.*, CS). For late fusion, we will first perform classification (*e.g.*, via CS) based on each modality separately, and then combine/average their predictions to achieve the final diagnosis results.

Results Using Single Modality. As shown in Fig. 3, we report the disease classification results achieved by three machine learning methods based on single modality data. From Fig. 3, we have the following interesting observations. *First*, the GCN method is superior to other traditional methods when using functional connectivity (FC) features. It achieves at least 4% improvement compared with other methods. This demonstrates the necessity and effectiveness of exploiting graph topology on them. *Besides*, for two traditional learning methods and the GCN method, results on FC calculated from fMRI are slightly better than those on fractional anisotropy (FA) data derived from dMRI in most cases.

Results Using Multiple Modalities. To evaluate the efficiency of the proposed heterogeneous graph framework when dealing with multi-modality data, we compared our HGACN with three competing methods using two multi-

Table 2. Results of our HGACN and three competing methods (*i.e.*, CS, PS, and GCN) with early fusion and late fusion strategies in the task of ADHD vs. HC classification. The terms "EF_CS" and "LF_CS" denote CS with the early fusion and late fusion strategy, respectively.

Metric	EF_CS	LF_CS	EF_PS	LF_PS	EF_GCN	LF_GCN	HGACN
ACC	0.625(0.027)	0.631(0.031)	0.628(0.029)	0.622(0.049)	0.679(0.034)	0.667(0.042)	**0.701**(0.035)
SEN	0.583(0.030)	0.698(0.031)	0.743(0.052)	0.727(0.028)	0.669(0.038)	0.676(0.021)	**0.763**(0.045)
SPE	0.667(0.033)	0.574(0.029)	0.459(0.057)	0.467(0.048)	0.638(0.023)	**0.683**(0.031)	0.648(0.026)

modality fusion strategies, with results reported in Table 2. From Table 2, we can see that our HGACN method achieves at least 2% improvement in terms of ACC and SEN values, compared with the competing methods. This demonstrates the effectiveness of our proposed HGACN in integrating multimodal MRIs for ADHD diagnosis. Besides, with CS as the final discriminator, we can obtain better results using the late fusion strategy. In the remaining cases, early fusion is a better option. Recalling the results in Fig. 3, we can observe that one can obtain the overall better results using multi-modality data (no matter with early or late fusion).

These results imply that fMRI and sMRI contain complementary information that help boost the identification performance for ADHD patients.

4 Conclusion and Future Work

In this paper, we propose a heterogeneous graph attention convolution network (HGACN) to integrate different modalities data in a unified GCN framework. Specifically, we first calculate FC and FA features from fMRI and dMRI data. Then, each subject will construct a heterogeneous graph with two types of nodes and three different categories of edges. We further employ attention mechanisms to aggregate information between intra- and inter-metapaths. Finally, we use new graph embedding to predict the label of the subject. Thus, the proposed HGACN can effectively capture complementary information between different types of brain connectivity features. Experimental results on 187 subjects demonstrate that our method yields state-of-the-art performance in identifying adult ADHD patients from healthy controls.

In the current work, we only focus on functional/structural connectivity matrices to capture subject-level connectivity topology. Actually, information conveyed in structure MRIs such as tissue volume can help uncover the neurobiological mechanisms of ADHD, which will be considered in our future work. Also, we simply combine multimodal MRIs acquired from two imaging site in the current work and ignore their inter-site differences. In the future, we will focus on alleviating inter-site data heterogeneity to further boost ADHD identification.

References

1. Swanson, J.M., et al.: Attention deficit hyperactivity disorder. In: Encyclopedia of Cognitive Science (2006)
2. Milham, M.P., Fair, D., Mennes, M., Mostofsky, S.H., et al.: The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. Front. Syst. Neurosci. **6**, 62 (2012)
3. Kessler, R.C., et al.: The prevalence and correlates of adult ADHD in the United States: results from the National Comorbidity Survey Replication. Am. J. Psychiatry **163**(4), 716–723 (2006)
4. Kooij, S.J., et al.: European consensus statement on diagnosis and treatment of adult ADHD: the European Network Adult ADHD. BMC Psychiatry **10**(1), 1–24 (2010)
5. Tenev, A., Markovska-Simoska, S., Kocarev, L., Pop-Jordanov, J., Müller, A., Candrian, G.: Machine learning approach for classification of ADHD adults. Int. J. Psychophysiol. **93**(1), 162–166 (2014)
6. Yao, D., Sun, H., Guo, X., Calhoun, V.D., Sun, L., Sui, J.: ADHD classification within and cross cohort using an ensembled feature selection framework. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1265–1269. IEEE (2019)
7. Zhang, L., Wang, M., Liu, M., Zhang, D.: A survey on deep learning for neuroimaging-based brain disorder analysis. Front. Neurosci. **14**, 779 (2020)
8. Yang, E., et al.: Deep Bayesian hashing with center prior for multi-modal neuroimage retrieval. IEEE Trans. Med. Imaging **40**, 503–513 (2020)
9. Yao, D., et al.: A mutual multi-scale triplet graph convolutional network for classification of brain disorders using functional or structural connectivity. IEEE Trans. Med. Imaging **40**(4), 1279–1289 (2021)
10. Liu, M., Zhang, J., Yap, P.T., Shen, D.: View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data. Med. Image Anal. **36**, 123–134 (2017)
11. Yao, D., et al.: Discriminating ADHD from healthy controls using a novel feature selection method based on relative importance and ensemble learning. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4632–4635. IEEE (2018)
12. Soussia, M., Rekik, I.: Unsupervised manifold learning using high-order morphological brain networks derived from T1-w MRI for autism diagnosis. Front. Neuroinform. **12**, 70 (2018)
13. Mahjoub, I., Mahjoub, M.A., Rekik, I.: Brain multiplexes reveal morphological connectional biomarkers fingerprinting late brain dementia states. Sci. Rep. **8**(1), 1–14 (2018)
14. Cortese, S., et al.: Toward systems neuroscience of ADHD: a meta-analysis of 55 fMRI studies. Am. J. Psychiatry **169**(10), 1038–1055 (2012)
15. Guo, X., et al.: Shared and distinct resting functional connectivity in children and adults with attention-deficit/hyperactivity disorder. Transl. Psychiatry **10**(1), 1–12 (2020)
16. Ellison-Wright, I., Ellison-Wright, Z., Bullmore, E.: Structural brain change in attention deficit hyperactivity disorder identified by meta-analysis. BMC Psychiatry **8**(1), 51 (2008)
17. Xia, S., Li, X., Kimball, A.E., Kelly, M.S., Lesser, I., Branch, C.: Thalamic shape and connectivity abnormalities in children with attention-deficit/hyperactivity disorder. Psychiatry Res.: Neuroimaging **204**(2–3), 161–167 (2012)

18. Luo, Y., Alvarez, T.L., Halperin, J.M., Li, X.: Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset ADHD using ensemble learning techniques. NeuroImage: Clin **26**, 102238 (2020)
19. Ktena, S.I., et al.: Metric learning with spectral graph convolutions on brain connectivity networks. Neuroimage **169**, 431–442 (2018)
20. Ma, G., et al.: Deep graph similarity learning for brain data analysis. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 2743–2751 (2019)
21. Bessadok, A., Mahjoub, M.A., Rekik, I.: Graph neural networks in network neuroscience. arXiv preprint arXiv:2106.03535 (2021)
22. Xing, X., et al.: Dynamic spectral graph convolution networks with assistant task training for early MCI diagnosis. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11767, pp. 639–646. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32251-9_70
23. Song, X., Frangi, A., Xiao, X., Cao, J., Wang, T., Lei, B.: Integrating similarity awareness and adaptive calibration in graph convolution network to predict disease. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12267, pp. 124–133. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_13
24. Yu, S., et al.: Multi-scale enhanced graph convolutional network for early mild cognitive impairment detection. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12267, pp. 228–237. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_23
25. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)